

An overview of quantum error mitigation formulas

Dayue Qin(秦大粵), Xiaosi Xu(徐晓思), and Ying Li(李颖)
Citation: Chin. Phys. B, 2022, 31 (**9**): 090306. DOI: 10.1088/1674-1056/ac7b1e
Journal homepage: http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

Steering quantum nonlocalities of quantum dot system suffering from decoherence

Huan Yang(杨欢), Ling-Ling Xing(邢玲玲), Zhi-Yong Ding(丁智勇), Gang Zhang(张刚), and Liu Ye(叶柳)
Chin. Phys. B, 2022, 31 (**9**): 090302. DOI: 10.1088/1674-1056/ac615a

Exploring Majorana zero modes in iron-based superconductors

Geng Li(李更), Shiyu Zhu(朱诗雨), Peng Fan(范朋), Lu Cao(曹路), and Hong-Jun Gao(高鸿钧)
Chin. Phys. B, 2022, 31 (**8**): 080301. DOI: 10.1088/1674-1056/ac70c3

Quantum walk search algorithm for multi-objective searching with iteration auto-controlling on hypercube

Yao-Yao Jiang(姜瑶瑶), Peng-Cheng Chu(初鹏程), Wen-Bin Zhang(张文彬), and Hong-Yang Ma(马鸿洋)
Chin. Phys. B, 2022, 31 (**4**): 040307. DOI: 10.1088/1674-1056/ac2806

Vacuum-gap-based lumped element Josephson parametric amplifier

Sishi Wu(吴思诗), Dengke Zhang(张登科), Rui Wang(王锐), Yulong Liu(刘玉龙), Shuai-Peng Wang(王帅鹏), Qichun Liu(刘其春), J S Tsai(蔡兆申), and Tiefu Li(李铁夫)
Chin. Phys. B, 2022, 31 (**1**): 010306. DOI: 10.1088/1674-1056/ac280c

Low-loss belief propagation decoder with Tanner graph in quantum error-correction codes

Dan-Dan Yan(颜丹丹), Xing-Kui Fan(范兴奎), Zhen-Yu Chen(陈祯羽), and Hong-Yang Ma(马鸿洋)
Chin. Phys. B, 2022, 31 (**1**): 010304. DOI: 10.1088/1674-1056/ac11cf

---

TOPICAL REVIEW — Celebrating 30 Years of *Chinese Physics B*

An overview of quantum error mitigation formulas

Dayue Qin(秦大粵), Xiaosi Xu(徐晓思), and Ying Li(李颖)†

Graduate School of China Academy of Engineering Physics, Beijing 100193, *China*

(Received 24 April 2022; revised manuscript received 21 June 2022; accepted manuscript online 22 June 2022)

Minimizing the effect of noise is essential for quantum computers. The conventional method to protect qubits against noise is through quantum error correction. However, for current quantum hardware in the so-called noisy intermediate-scale quantum (NISQ) era, noise presents in these systems and is too high for error correction to be beneficial. Quantum error mitigation is a set of alternative methods for minimizing errors, including error extrapolation, probabilistic error cancellation, measurement error mitigation, subspace expansion, symmetry verification, virtual distillation, etc. The requirement for these methods is usually less demanding than error correction. Quantum error mitigation is a promising way of reducing errors on NISQ quantum computers. This paper gives a comprehensive introduction to quantum error mitigation. The state-of-art error mitigation methods are covered and formulated in a general form, which provides a basis for comparing, combining and optimizing different methods in future work.

Keywords: quantum error mitigation, quantum computing, quantum error correction, noisy intermediate-scale quantum

PACS: 03.67.Pp **DOI:** 10.1088/1674-1056/ac7b1e

1. Introduction

Quantum computers have the potential of outperforming classical computers in solving certain problems.[1] With a particular problem, one needs to map a quantum algorithm into quantum circuits that the quantum hardware can execute. However, the quantum state stored in any physical system, no matter superconducting qubits, trapped ions, etc.,[2] has a limited life time due to interactions with the environment. Besides, any operation is associated with noise. These altogether lead to unreliable computing outcomes. For superconducting qubits, the error rate per two-qubit gate, i.e., the probability of an error occurring in the gate, is about 0.5% with current technology.[3] The error rate limits the number of quantum gates in a circuit, to merely a thousand.[4,5] Therefore, it is necessary to minimize the impact of noise in a quantum circuit, generally by the means of quantum error correction (QEC) or quantum error mitigation (QEM).

QEC is the conventional method to correct errors by taking the use of quantum error correction codes. Quantum states are encoded into some code space, which is a subspace of the entire Hilbert space – physically, a logical qubit is encoded into several physical qubits. The encoding, error detection and correction are realized with quantum gates that are also noise-prone. Therefore, if the physical error rate per gate is higher than a certain threshold, performing error correction can introduce more errors than that it can correct.[6] QEC also requires an overhead cost on qubits for encoding the logical information. For example, when the physical error rate is 10^{-3}, about 1.4 million qubits are required to calculate the electronic structure of 54 spin-orbitals using the surface code.[7] With the current technology, it is still challenging to experimentally demonstrate beneficial error correction in quantum computing.

As the name suggests, QEM methods are designed for mitigating the effect of errors instead of correcting them. They can work in a quantum circuit with an error rate higher than the threshold of QEC, without the large qubit overhead. Therefore, QEM is suitable for quantum devices in the noisy intermediate-scale quantum (NISQ) era,[8] a midway towards fault-tolerant quantum computers[9] in which quantum computing is fully protected by QEC. NISQ devices have only dozens to hundreds of qubits, nevertheless, they have demonstrated the quantum advantage on a specific problem,[4,10] and it is believed that they can retain this advantage to solve practical problems. To achieve this goal, many quantum algorithms[11,12] have been proposed. Incorporating QEM methods is essential for certain applications on NISQ devices.

Various QEM methods have been proposed. In methods such as error extrapolation (or zero-noise extrapolation)[13,14] and probabilistic error cancellation (PEC),[14,15] given some knowledge about the noise, we can compensate errors and make the computing result closer to the true value. These methods do not require additional qubits. The knowledge of the noise, e.g., a parameter such as the error rate or a concrete error model, can be obtained with some characterization methods.[16,17] Alternatively, the parameters used in QEM protocols can be optimized in a learning-based or data-driven manner.[18,19] In symmetry verification[20,21] and vir-

†Corresponding author. E-mail: yli@gscaep.ac.cn

© 2022 Chinese Physical Society and IOP Publishing Ltd

tual distillation,[22–25] errors are eliminated based on certain constrains on the quantum state, i.e., symmetry and purity, respectively. Multiple copies of the noisy state are required for effectively increasing the purity in virtual distillation. These methods are more similar to QEC than error extrapolation and PEC: In a QEC code, a symmetry (i.e., the stabilizer group in stabilizer codes) is constructed at the cost of qubit overhead. In addition to general-purpose QEM methods, protocols are also developed for specific algorithms.[26–28] For example, the quantum subspace expansion[26,27] is a protocol designed to reduce errors in a variational quantum eigensolver.[29]

In this paper, we summarize the state-of-the-art error mitigation methods. Although they work in different ways, we formulate them in a general form. In this way, we explicitly define two characteristic quantities of all QEM protocols, the bias and the variance. With the general form, there is a universal way to combine different protocols and optimize them in the leaning-based manner. In Section 2, we introduce the notations and the general form. QEM methods are specifically described and discussed from Section 3 to Section 7. We introduce the Clifford-circuit learning method in Section 8 and summarize the combined methods in Section 9. A summary is given in Section 10.

2. Notations

Here we introduce the notations used throughout this paper. We consider a quantum circuit with n qubits and N gates. We always assume that qubits are prepared in $|0\rangle^{\otimes n}$. When free of errors, the quantum circuit realizes the unitary operator $U = U_N U_{N-1} \cdots U_1$, where U_i is the error-free operator of the i-th gate. At the end of the quantum circuit, the expected value of an observable O is measured. We use $C = (U_N, U_{N-1}, \ldots U_1, O)$ to denote the quantum circuit with a specified observable. The error-free final state and computation result are respectively expressed as $\rho_C^{\rm ef} = [U](|0\rangle\langle 0|^{\otimes n})$ and $f_C = {\rm Tr}\left[O\rho_C^{\rm ef}\right]$, where $[U](\cdot) \equiv U \cdot U^\dagger$. When subjected to noise, the operations of quantum gates are changed to a series of complete-positive maps $\mathcal{U} = \mathcal{U}_N \circ \mathcal{U}_{N-1} \circ \cdots \circ \mathcal{U}_1$, where $\mathcal{U}_i \circ \mathcal{U}_j(\cdot) \equiv \mathcal{U}_i(\mathcal{U}_j(\cdot))$. Typically we can express $\mathcal{U}_i = \mathcal{E}_i \circ [U_i]$, where $\mathcal{E}_i$ is the noise operation associated with the i-th gate. The erroneous final state and computation result are respectively expressed as $\rho_C = \mathcal{U}(|0\rangle\langle 0|^{\otimes n})$ and $y_C = {\rm Tr}[O\rho_C]$.

Generally speaking, provided with the original circuit C, quantum error mitigation performs a set of variant circuits $(C_1, C_2, \ldots)$ with noisy operations and then obtains the error-mitigated result from a formula F. We use f_C' to denote the error-mitigated result. The general form of the error-mitigated result is

$$f_C' = F(y_{C_1}, y_{C_2}, \ldots, \lambda_1, \lambda_2, \ldots), \tag{1}$$

where $(\lambda_1, \lambda_2, \ldots)$ are the parameters. The formula F as well as the choice of $(C_1, C_2, \ldots)$ depend on the error-mitigation method. Different error mitigation methods employ different techniques to determine the parameters in the formula.

We characterize an error mitigation method with two quantities, the bias and the variance. In quantum computing, y_C is estimated with the random outputs of the circuit C. We denote the estimator of y_C with $\hat{y}_C$. Usually, $\hat{y}_C$ is unbiased, i.e., $\mathrm{E}[\hat{y}_C] = y_C$, and the variance depends on the number of circuit executions. For example, if the observable $O = Z$ is the Pauli operator of a qubit, the measurement outcome of one circuit execution is a random number ± 1. According to the empirical mean estimator, $\hat{y}_C$ is the average value of measurement outcomes, and its variance is $(1 - y_C^2)/M$, where M is the number of circuit executions. Given $\hat{y}_C$, the estimator of f_C' is $\hat{f}_C' = F(\hat{y}_{C_1}, \hat{y}_{C_2}, \ldots)$ according to the error mitigation formula. Bias and variance of the QEM estimator are respectively $\mathrm{E}[\hat{f}_C'] - f_C$ and $\mathrm{E}[\hat{f}_C'^2] - \mathrm{E}[\hat{f}_C']^2$. In some error extrapolation protocols and PEC, the formula is linear in y_C, thus $\mathrm{E}[\hat{f}_C'] = f_C'$ if each $\hat{y}_C$ is unbiased, and the bias of the QEM estimator is $f_C' - f_C$; in general this result on the bias does not hold. Usually, error mitigation can reduce the bias while increasing the variance, in comparison with the raw quantum computing without error mitigation (corresponding to $f_C' = y_C$). The schematic diagram of error-free, erroneous and error mitigated computation results can be found in Fig. 1.

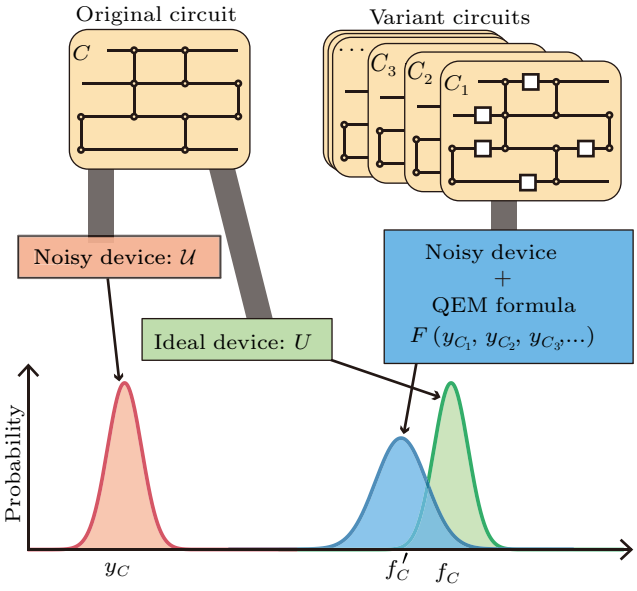


Fig. 1. Schematic diagram of quantum computing with and without QEM. The output of the original circuit is f_C on an ideal quantum device and y_C on a noisy device. Quantum error mitigation performs computations on a set of variant circuits to predict the error-free result with an increased accuracy. A general cost of quantum error mitigation is the enlarged variance.

3. Error extrapolation

In error extrapolation, the basic routine is to execute the quantum circuit at various noise levels and then extrapolate to the zero-noise limit. The two essential components are extrapolation formulas and error-boosting methods. Error-boosting methods increase the error rate while preserving the

error model. With the computation results at increased error rates, we can predict the error-free result according to the extrapolation formulas.

3.1. Polynomial extrapolation

The idea of error extrapolation was introduced in works by Li and benjamin[13] and Temme *et al.*,[14] in which the polynomial fitting function is considered. It is assumed that the erroneous result can be expressed with the polynomial decomposition

$$y_C(\epsilon) = f_C + \sum_{k=1}^{m} a_k \epsilon^k + O(\epsilon^{m+1}), \tag{2}$$

where ϵ is the total error rate of gates in the circuit. Executing the quantum circuit with l $(\geq m+1)$ different error rates, we can eliminate the first m-th order effects of noise using an m-order polynomial fitting.

Richardson extrapolation When $l = m+1$, we can employ the Richardson extrapolation formula

$$f'_C = \sum_{j=0}^{m} \lambda_j y_C(\epsilon_j), \tag{3}$$

where ϵ_j is the j-th error rate and the parameters are the solution of the following equations:

$$\sum_{j=0}^{m} \lambda_j = 1, \ \sum_{j=0}^{m} \lambda_j \epsilon_j^k = 0, \ \forall k = 1, 2, \ldots, m. \tag{4}$$

Truncated Neumann series Recently, Wang *et al.*[30] proposed error mitigation with truncated Neumann series. In this protocol, a formula resembling that in Eq. (3) is used, but the parameters are determined according to the truncated Neumann series. Let $F : \mathcal{E} \to \mathbb{R}$ be a linear map from a superoperator $\mathcal{E}$, which is a squared matrix in the Pauli transfer matrix representation,[17] to the real field $\mathbb{R}$. Under the condition of $\lim_{k\to\infty}(\mathbb{I}-\mathcal{E})^k = 0$, which can be satisfied if $\|\mathbb{I}-\mathcal{E}\| < 1$ in some consistent norm, it holds that

$$F(\mathbb{I}) = \sum_{k=1}^{m+1} \lambda_k F(\mathcal{E}^k) + F((\mathbb{I}-\mathcal{E})^{m+1}), \tag{5}$$

where $\lambda_k = (-1)^{k-1}\binom{m+1}{k}$, and $\binom{m+1}{k}$ is the binomial coefficient. The proof of Eq. (5) can be found in Appendix B of Ref. [30]. Suppose $\mathcal{U} = \mathcal{E} \circ [U]$, where $\mathcal{E}$ is the overall noise effect and $[U]$ prepares the error-free state ρ_C^{ef}. Replacing $F(\mathbb{I})$ with $f_C = \text{Tr}(O[\mathbb{I}](\rho_C^{\text{ef}}))$ and $F(\mathcal{E}^k)$ with $y_C(\mathcal{E}^k) = \text{Tr}(O\mathcal{E}^k(\rho_C^{\text{ef}}))$ in Eq. (5), we have

$$f'_C = \sum_{k=1}^{m+1} \lambda_k y_C(\mathcal{E}^k). \tag{6}$$

Here $y_C(\mathcal{E}^k)$ can be realized by repeatedly applying $\mathcal{E}$ for k times. The bias in Eq. (6) decays exponentially with k when the error rate is small.

3.2. Exponential extrapolation

In some cases, the exponential extrapolation proposed by Endo *et al.*[15] has advantages in accuracy. The simplest exponential extrapolation is

$$f'_C = y_C(\epsilon_0)^{\frac{\epsilon_1}{\epsilon_1-\epsilon_0}} y_C(\epsilon_1)^{\frac{\epsilon_0}{\epsilon_0-\epsilon_1}}, \tag{7}$$

given two error rates $\epsilon_0 < \epsilon_1$. Giurgica-Tiron *et al.*[31] summarized the extrapolation formulas and introduced a more general form of poly-exponential extrapolation, in which the logarithm of erroneous result fits to a polynomial function. Cai[32] generalized error extrapolation to the multi-exponential functions where the sum of exponents is the fitting function.

The advantage of exponential extrapolation is analyzed by Endo *et al.*[15] and Cai.[32] Here we briefly summarize the results. In the circuit model, the number of locations prone to errors equals to the number of gates. With a sufficiently large number of gates, the probability of errors occurring at l locations satisfies the Poisson distribution $P(l) = \text{e}^{-\mu}\frac{(\mu)^l}{l!}$, where μ is the average number of error-occurred locations. Assuming that the occurred error will move the measured expectation towards zero with a probability of α, the erroneous result is

$$y_C = \sum_{l=0}^{N} f_C(1-\alpha)^l \text{e}^{-\mu}\frac{(\mu)^l}{l!}, \tag{8}$$

and in the $N \to \infty$ limit it becomes an exponential function $f_C\text{e}^{-\mu\alpha}$. The advantage of exponential extrapolation is numerically demonstrated in Refs. [15,31]. Observations of the fidelity decay in experiments, e.g., the results in Ref. [33], also support this conclusion.

3.3. Resource cost

Here we briefly discuss the resource cost of error extrapolation. We take Richardson extrapolation as an example. One may think that we can arbitrarily improve the precision by increasing the extrapolation order m in Eq. (2). This is true only when we can exactly measure the value of $y_C(\epsilon)$, which is impossible because of the shot noise. In fact, if the extrapolation curve interpolates all the data points, such as that in Richardson extrapolation, the variance is increased by a factor around $\sum_{j=0}^{m}\lambda_j^2$. Therefore, we should in practice choose a proper extrapolation order m. Krebsbach *et al.*[34] studied the variance–bias trade-off in Richardson extrapolation and proposed the optimal extrapolation order m, spacing $|\epsilon_j - \epsilon_{j-1}|$ for each noise rate given the overall sampling overhead. From the perspective of statistics, the increased variance is referred to the phenomenon of overfitting,[35] which can be suppressed with increased number of data points l $(\gg m)$ and by using fitting methods such as least squares rather than Richardson extrapolation.

3.4. Error-boosting methods

The first method takes use of Pauli twirling and random Pauli gate insertion.[13] This method relies on the assumption that quantum errors of the single-qubit gates can be ignored comparing with those of multi-qubit gates. We can effectively increase the error via randomly inserting Pauli gates after each multi-qubit gate. Pauli twirling converts an arbitrary noise channel $\mathcal{E}_i$ into Pauli noise:[36]

$$\mathcal{E}_i^{\rm tr} = \sum_{ij} P_{i,j}[\sigma_i\sigma_j], \tag{9}$$

where $\sigma_i, \sigma_j \in \{\sigma_I, \sigma_x, \sigma_y, \sigma_z\}$ are single-qubit Pauli operators and $\sigma_i\sigma_j$ denotes a tensor product. P_{ij} are positive coefficients satisfying $\sum_{i,j} P_{ij} = 1$. To implement the Pauli twirling technique, we randomly insert Pauli gates before and after $\mathcal{U}_i$, resulting in $[\sigma_d\sigma_c]\circ\mathcal{U}_i\circ[\sigma_b\sigma_a]$, where σ_a and σ_b are uniformly sampled from $\{\sigma_I, \sigma_x, \sigma_y, \sigma_z\}$ and accordingly $\sigma_d\sigma_c = [U_i]\sigma_b\sigma_a$. To boost up the error rate, a further step is to randomly insert Pauli gates after each multi-qubit gate. Recall Eq. (9) and assume a small error rate, randomly inserting each Pauli gate $\sigma_i\sigma_j$ with a probability of $rP_{i,j}$ can boost the error rate from ϵ to $(1+r)\epsilon$.

The second method is to insert identity-equivalent gates. The approaches to this end include repeating self-invert gates for odd times,[37,38] gate folding[31] and identity insertion.[39] Typically, this method is based on

$$\tilde{U} = \tilde{U}(\tilde{U}^\dagger\tilde{U})^r, \tag{10}$$

for arbitrary unitary operator $\tilde{U}$ and non-negative integer r. Gate folding according to the above equation will amplify the error rate by a factor of $1+2r$ if the noise operation commutes with the unitary operator and the error rate is small. To be more detailed, suppose the noise operation on the i-th gate is in the form of $\mathcal{E}_i = (1-\alpha)[\mathbb{I}] + \alpha\mathcal{N}_i$, the noise operation on the folded gates is $\mathcal{E}_i^r = (1-r'\alpha)[\mathbb{I}] + r'\alpha\mathcal{N}_i + O(\alpha^2)$, where $r' = 1+2r$. Here we have assumed $\mathcal{N}_i$ commutes with $\tilde{U}$. In the case that α is small, the error rate is enlarged by a factor of $1+2r$. Primitively, we should fold every erroneous gate with the same r, such that the error model is unchanged. In that case, we can only amplify the total error rate for $1,3,5,\ldots$ times. To achieve denser spectrum of error rates, we can choose to fold m out of N gates and the error rate is increased by a factor of $1+2rm/N$. Be aware of that partially folding a circuit may not only increase the total error rate but also severely change the error model, thus it should be subtle when choosing which gate to be folded. To keep the error model as unchanged as possible, Giurgica-Tiron *et al.*[31] divided the quantum gates into subsets and chose to fold particular gates in each of the subsets. He *et al.*[39] used the method of random identity insertion. They set r of each gate as a random variable and sampled circuits according to the distribution of r. Recently, Schultz *et al.*[40] introduced the method of gate Trotterization. They proposed compiling a gate G into products of several gates according to the formula $G = (G^{1/r})^r$.

The third method is stretching the Hamiltonian. On realistic quantum hardware, quantum gates are implemented by evolving qubits' state under a controlled time-dependent Hamiltonian denoted by $H(t)$. The stretched Hamiltonian $H(t/r)/r$ can increase the error rate by a multiplicative factor of r under the assumption that the noise generator is independent of the Hamiltonian and constant in time.[14] In experiments, we can realize the stretched Hamiltonian by reducing the pulse strength that controls the qubits while prolonging the pulse time. An illustrative example is, similar to the Rabi stretch for single-qubit in Ref. [41], a spin initialized in the Z-direction evolving under a X-direction magnetic field $H = B\sigma_x(t)$. The method of stretching the Hamiltonian here is to increase the duration t and reduce the field strength B by the same multiple. A more practical example is given in Ref. [42], where it tunes the cross-resonance pulse on a superconducting quantum processor.

Here we briefly discuss the advantages and disadvantages of these three error-boosting methods. The first method requires accurate understanding of the error model, and the error rate is boosted in a pseudo manner, i.e., the error rate is not actually increased but the collective result of random circuits is equal to the result on an error-increased device. The advantage is that, with the aid of reliable gate benchmarking methods such as gate set tomography,[17] the critical condition that boosting error will not change the error model can be fulfilled. The second method is easy to implement. Comparing with the first one, the second method does not need any definite knowledge about the noise model. However, it is at greater risk that the error model is changed when folding gates. This is not a serious problem when with only simple error models such as depolarizing. In general, it may cause that, quoted from Ref. [5], "arbitrarily deviate from the desired result when only slightly more complicated error models are considered". Both the first and second methods are in the digital level and friendly to cloud-platform users, while the third method requires direct control of the underlying quantum hardware. Nevertheless, pulse-level control is accessible through some hardware providers.[43,44] As shown in the work of Schultz *et al.*,[40] the amount of increased noise by different methods depends on the time-correlation. The performance of different methods is similar with white noise, but can be significantly different for time-correlated noise.

3.5. Multiple noise parameter extrapolation

Hyper-surface fitting[45] and error mitigation via individual error reduction[46] are proposed by Otten and Gray. Both

works considered the linear error mitigation formula with multiple noise parameters.

Hyper-surface fitting In reality, a quantum circuit may be subjected to different noise processes, e.g., relaxation and dephasing, thus is characterized with separate error parameters. Suppose there are m distinct error parameters $\epsilon_{(1)},\epsilon_{(2)},\ldots,\epsilon_{(m)}$, we can generalize Eq. (2) to a hyper-surface:

$$y_C\left(\epsilon_{(1)},\epsilon_{(2)},\ldots,\epsilon_{(m)}\right)=f_C+\sum_{i=1}^{m}a_i\epsilon_{(i)}+\sum_{i,j=1}^{m}a_{ij}\epsilon_{(i)}\epsilon_{(j)}+\cdots. \tag{11}$$

We can use the method of least squares to fit this hyper-surface.[45] For example, if there are two error parameters $(\epsilon_{(1)},\epsilon_{(2)})=(\epsilon,\delta)$, we can truncate Eq. (11) to the second order and solve

$$EX=Y, \tag{12}$$

where

$$\begin{aligned}E&=\begin{pmatrix}1&\epsilon_1&\delta_1&\epsilon_1^2&\delta_1^2&\epsilon_1\delta_1\\1&\epsilon_2&\delta_2&\epsilon_2^2&\delta_2^2&\epsilon_2\delta_2\\1&\epsilon_3&\delta_3&\epsilon_3^2&\delta_3^2&\epsilon_3\delta_3\\\vdots&\vdots&\vdots&\vdots&\vdots&\vdots\end{pmatrix},\\X&=(\widetilde{f_C},a_1,a_2,a_{11},a_{22},a_{12})^{\mathrm{T}},\\Y&=(y_C(\epsilon_1,\delta_1),y_C(\epsilon_2,\delta_2),y_C(\epsilon_3,\delta_3),\ldots)^{\mathrm{T}}.\end{aligned}$$

By minimizing the difference between the left- and right-hand sides of this equation, the error mitigation formula is

$$f_C'=\arg\min_{\widetilde{f_C}}\left(\min_{A}||EX-Y||^2\right), \tag{13}$$

where $A=(a_1,a_2,a_{11},a_{22},a_{12})^{\mathrm{T}}$ and $||\cdot||$ is the Euclidean norm.

Individual error reduction We may divide a quantum computer into several subsystems, e.g., qubits, and each of them can be considered has an individual error source. We write the noise operation $\mathcal{E}_i$ in the Lindblad form $\mathcal{E}_i=\exp(\tau\sum_{j=1}^{m}L_j)$, where τ is the gate time and $L_1,L_2,\ldots,L_m$ are Lindblad operators accounting for errors in each of the m subsystems. The first order approximation of $\mathcal{E}_i$ is

$$\mathcal{E}_i\approx 1+\tau\sum_{j=1}^{m}L_j. \tag{14}$$

Even though full fault tolerance is prohibitive on a NISQ device, error correction for individual error source may be attainable. With error correction, suppose that we can have the k-th error component reduced, $\mathcal{E}_i$ becomes

$$\mathcal{E}_i^{(k)}\approx 1+\tau\left(\sum_{j\neq k}^{m}L_j+(1-r_k)L_k\right), \tag{15}$$

where r_k quantifies the removed fraction. We consider the density matrix defined as

$$\tilde{\rho}=\rho-\sum_{k=1}^{m}\frac{1}{r_k}\left(\rho-\rho^{(k)}\right), \tag{16}$$

where ρ is the final quantum state prepared by $\mathcal{U}$ and $\rho^{(k)}$ is the final quantum state prepared by $\mathcal{U}^{(k)}=\mathcal{E}_N^{(k)}[U_N]\mathcal{E}_{N-1}^{(k)}[U_{N-1}]\cdots\mathcal{E}_1^{(k)}[U_1]$. It follows that

$$\tilde{\rho}=\rho^{\mathrm{ef}}+O(\tau^2), \tag{17}$$

where ρ^{ef} is the error-free state. The proof of Eq. (17) can be found in the supporting material of Ref. [46]. This immediately leads to the error-mitigation formula

$$f_C'=y_C-\sum_{k=1}^{m}\frac{1}{r_k}\left(y_C-y_C^{(k)}\right), \tag{18}$$

where $y_C^{(k)}=\mathrm{Tr}[O\mathcal{U}^{(k)}(|0\rangle\langle 0|^n)]$. Note that the additional qubits used to perform individual error correction may introduce extra errors and change the error model of the whole circuit.

4. Probabilistic error cancellation

Probabilistic error cancellation (PEC) was proposed by Temme *et al.*,[14] which is implemented based on a quasi-probability decomposition. A practical method of constructing the formula using gate set tomography and universal gate set was introduced by Endo *et al.*[15] The basic idea is that an error-free operation can be expressed as a linear combination of erroneous ones. Suppose that the vector $\boldsymbol{\beta}=(\beta_1,\beta_2,\ldots,\beta_N)$ specifies a quantum circuit (β_i specifies the i-th gate) and there exists a circuit decomposition $[U_{\boldsymbol{\beta}}]=\sum_{\boldsymbol{\beta}}\lambda_{\boldsymbol{\beta}}\mathcal{U}_{\boldsymbol{\beta}}$, the error cancellation formula is

$$f_C'=\sum_{\boldsymbol{\beta}}\lambda_{\boldsymbol{\beta}}y_C(\boldsymbol{\beta}), \tag{19}$$

where $y_C(\boldsymbol{\beta})=\mathrm{Tr}[O\mathcal{U}_{\boldsymbol{\beta}}(|0\rangle\langle 0|^n)]$.

One method to determine the coefficients $\lambda_{\boldsymbol{\beta}}$ is to find the quasi-probability decomposition for every gate. With the gate-wise decomposition

$$[U_i]=\sum_{\beta_i}\lambda_{i,\beta_i}\mathcal{U}_{i,\beta_i}, \tag{20}$$

where $\lambda_{i,\beta_i}\in\mathbb{R}$ and $\sum_{\beta_i}\lambda_{i,\beta_i}=1$ due to the trace-preserving condition. Terms in the circuit decomposition are $\mathcal{U}_{\boldsymbol{\beta}}=\mathcal{U}_{N,\beta_N}\circ\mathcal{U}_{N-1,\beta_{N-1}}\circ\cdots\circ\mathcal{U}_{1,\beta_1}$, and $\lambda_{\boldsymbol{\beta}}=\prod_i\lambda_{i,\beta_i}$. To implement the error cancellation, we need to use the Monte Carlo method. We rewrite Eq. (19) as

$$f_C'=\gamma\sum_{\boldsymbol{\beta}}\frac{|\lambda_{\boldsymbol{\beta}}|}{\gamma}\mathrm{sign}(\lambda_{\boldsymbol{\beta}})y_C(\boldsymbol{\beta}), \tag{21}$$

where $\gamma=\sum_{\boldsymbol{\beta}}|\lambda_{\boldsymbol{\beta}}|=\prod_{i=1}^{N}\gamma_i$ and $\gamma_i=\sum_{\beta_i}|\lambda_{i,\beta_i}|$. The error-mitigated result f_C' can be estimated by sampling circuits according to the probability $|\lambda_{\boldsymbol{\beta}}|/\gamma$.

One method to determine the gate-wise decomposition is to find the inverse of the quantum noise. Here we take the inverse of Pauli noise as an example. Suppose that the i-th gate is a single-qubit gate subjected to the Pauli noise $\mathcal{E}_i = \sum_{\beta_i} P_{i,\beta_i}[\sigma_{\beta_i}]$ where $\sigma_{\beta_i} \in \{\sigma_I, \sigma_x, \sigma_y, \sigma_z\}$ and P_{i,β_i} are probabilities, the coefficients of its inverse $\mathcal{E}_i^{-1} = \sum_{\beta_i} \lambda_{i,\beta_i}[\sigma_{\beta_i}]$ can be determined by solving

$$\begin{pmatrix} \lambda_{i,I} & \lambda_{i,x} & \lambda_{i,y} & \lambda_{i,z} \\ \lambda_{i,x} & \lambda_{i,I} & \lambda_{i,z} & \lambda_{i,y} \\ \lambda_{i,y} & \lambda_{i,z} & \lambda_{i,I} & \lambda_{i,x} \\ \lambda_{i,z} & \lambda_{i,y} & \lambda_{i,x} & \lambda_{i,I} \end{pmatrix} \begin{pmatrix} P_{i,I} & P_{i,x} & P_{i,y} & P_{i,z} \\ P_{i,x} & P_{i,I} & P_{i,z} & P_{i,y} \\ P_{i,y} & P_{i,z} & P_{i,I} & P_{i,x} \\ P_{i,z} & P_{i,y} & P_{i,x} & P_{i,I} \end{pmatrix} = \mathbb{I}, \tag{22}$$

where $\mathbb{I}$ is the 4×4 identity matrix. The method to construct $\mathcal{U}_{i,\beta_i}$ is by inserting Pauli gates σ_{β_i} after each $\mathcal{U}_i$, i.e., $\mathcal{U}_{i,\beta_i} = [\sigma_{\beta_i}] \circ \mathcal{E}_i \circ [U_i]$.

The inverse of typical error models, such as depolarizing and dephasing, can be decomposed with Pauli basis $\sigma_{\beta_i} \in \{\sigma_I, \sigma_x, \sigma_y, \sigma_z\}$. For a general error model, the decomposition basis is a larger gate set[15] whose elements can be synthesized with Clifford gates and the projection operator $\frac{1}{2}(\sigma_I + \sigma_z)$.

4.1. Resource cost

The sampling overhead of PEC is determined by γ. In comparison with the case without error mitigation, the variance in PEC is increased by a factor of γ^2: if we use Monte Carlo estimation according to Eq. (21), we need γ^2 times more circuit executions in order to achieve the same variance as in the error-free case. Recall that $\gamma = \prod_{i=1}^{N} \gamma_i$, the sampling overhead increases exponentially with the number of gates N. Consider $\gamma_i = 1 + O(\epsilon/N)$ and $\prod_{i=1}^{N}(1+\epsilon/N) \approx \mathrm{e}^{\epsilon}$ where ϵ is the total error rate, PEC is efficient when ϵ is small.

The sampling overhead depends on the choice of decomposition basis set. For a general noise channel, we can always decompose its inverse with the universal basis set[15] despite that it may not be optimal, i.e., there may exit other decomposition sets, with which we can achieve lower sampling overhead. One example given by Takagi[47] is the cancellation of the general dephasing noise $\mathcal{E} = (1-\epsilon)[\mathbb{I}] + \epsilon[\exp(-\mathrm{i}\hat{n}\cdot\hat{\sigma}\pi/n)]$ where $\hat{n}\cdot\hat{\sigma} = n_x\sigma_x + n_y\sigma_y + n_z\sigma_z$. It is shown that, when $\hat{n} = \left(\cos(\frac{\pi}{8}), 0, \sin(\frac{\pi}{8})\right)$, the sampling overhead is $\gamma = \frac{1+(\sqrt{2}-1)\epsilon}{1-2\epsilon}$ with the universal basis set, while the optimal overhead is $\gamma_{\mathrm{opt}} = 1/(1-2\epsilon)$ with the basis $\{[\mathbb{I}], \exp(\mathrm{i}\hat{n}\cdot\hat{\sigma}\pi/n)\}$. Indeed we can achieve the optimal overhead if we use a flexible decomposition set. One method to find the optimal decomposition set is by the method of constraint optimization.[14,48,49] Realistic quantum noise is complete-positive and trace-preserving (CPTP). As given in Ref. [49], the inverse of a CPTP map is Hermitian-preserving and trace-preserving (HPTP), and there exists the following decomposition:

$$\mathcal{E}^{-1} = \lambda_1\mathcal{C}_1 - \lambda_2\mathcal{C}_2, \tag{23}$$

where $\mathcal{E}^{-1}$ is a HPTP map, $\mathcal{C}_1, \mathcal{C}_2$ are CPTP maps, and $\lambda_1, \lambda_2 \geq 0$ are two real numbers. According to Eq. (23), we can employ the following constraint optimization to determine the optimal resource cost:

$$\gamma_{\mathrm{opt}} = \min(\lambda_1 + \lambda_2) \tag{24}$$

$$\text{subjected to: } J_{\mathcal{E}} = \lambda_1 J_{\mathcal{C}_1} - \lambda_2 J_{\mathcal{C}_2}, \tag{25}$$

$$\mathrm{Tr}_B(J_{\mathcal{C}_1}) = \mathrm{Tr}_B(J_{\mathcal{C}_2}) = \mathbb{I}_A, \tag{26}$$

$$J_{\mathcal{C}_1} \geq 1,\ J_{\mathcal{C}_2} \geq 1, \tag{27}$$

$$J_{\mathcal{C}_1} = J_{\mathcal{C}_1}^{\dagger},\ J_{\mathcal{C}_2} = J_{\mathcal{C}_2}^{\dagger}, \tag{28}$$

where $J_\Lambda = \sum_{i,j=0}^{d} |i\rangle\langle j|_A \otimes \Lambda(|i\rangle\langle j|_B)$ is the Choi-Jamiolkowski isomorphism of Λ and Tr_B denotes the partial trace over the subsystem B. Equation (26) comes from trace preservation, equation (27) comes from complete positivity and equation (28) comes from the Hermitian preservation of $\mathcal{E}^{-1}$. The physical implementation of a CPTP map such as $\mathcal{C}_1, \mathcal{C}_2$ can be realized on a quantum computer with at most extra n qubits according to the theorem of Stinespring dilation.[48]

4.2. Continuous-time formalism

Sun *et al.*[50] proposed the stochastic PEC, which extends the formulism of PEC for analogue quantum computation, such that the PEC protocol can be engineered at the pulse level and attain extra resilience against gate dependence and non-local noise effects. Stochastic here means randomly choosing time points to add correction operation instead of checking everywhere. We for now modify our notations to adjust to the continuous time formalism. Other than the total number of gates, we here consider $N = T/\delta t$, where δt is a small time interval and T is the time when the final state is prepared. During the i-th time interval, the time evolution on a realistic quantum machine is

$$\rho(t_i + \delta t) = \mathcal{E}(t_i)U(t_i)\rho(t), \tag{29}$$

where $\mathcal{E}(t_i)$ and $U(t_i)$ respectively correspond to $\mathcal{E}_i$ and U_i in the discrete-gate formalism. We consider the following decomposition of the inverse $\mathcal{E}^{-1}(t_i)$ on an operator set $\{\mathcal{B}_{ij}\}$:

$$\mathcal{E}^{-1}(t_i) = \gamma_i \left[p_0(t_i)\mathcal{I} + \sum_{j\geq 1} p_j(t_i)\mathcal{B}_{ij} \right], \tag{30}$$

where $\mathcal{I}$ is the identity evolution, $p_j(t_i)$ are positive probabilities, and $\mathcal{B}_{ij}$ has absorbed the sign. The coefficients $\gamma_i = 1 + O(\delta t)$, and $p_0(t_i) = 1 - \sum_{j\geq 1} p_j(t_i) = 1 - O(\delta t)$. In practice, it would be excessively time-consuming if we try to correct $\mathcal{E}(t_i)$ at every time step t_i. Considering the fact that, with $p_0 \lesssim 1$, we do nothing at most time points, i.e., the correction $\mathcal{B}_{ij}$ is placed at a small number of time points. Thus, one can alternately investigate the distribution of the time duration

between two corrections. The probability of no correction is added until time t is

$$P(t) = \lim_{\delta t \to 0} \prod_{i=0}^{t/\delta_t} \left[1 - \sum_{j \geq 1} p_j(t_i) \right] = \mathrm{e}^{-\int_0^t \Gamma(t')\,\mathrm{d}t'}, \tag{31}$$

where $\Gamma(t') = \sum_{j\geq 1} p_j(t_i)/\delta_t = O(1)$. The probability of placing a correction in the interval $(t, t+\delta t)$ is $\Gamma(t)\mathrm{d}t$. Thus the probability density of time duration between two corrections is $P(t)\Gamma(t)$, which can be sampled using the method of inverse transform sampling.

The overhead in the continuous time formalism is given by Hakoshima *et al.*[51] Rewriting Eq. (30) in the form of

$$\mathcal{E}^{-1}(t_i) = (1 + q_0(t_i))\mathcal{I} + \sum_{j\geq 1} q_j(t_i)\mathcal{B}_{ij}, \tag{32}$$

where the sign is absorbed in $q_j(t_j)$, we have

$$\begin{aligned} \gamma(T) &= \lim_{\delta t \to 0} \prod_{i=0}^{T/\delta t} \left[1 + q_0(t_i) + \sum_{j\geq 1} |q_j(t_i)| \right] \\ &= \exp\left[\int_0^T \left(\frac{q_0(t) + \sum_{j\geq 1} |q_j(t)|}{\delta t} \right) \mathrm{d}t \right]. \end{aligned} \tag{33}$$

The above equation enables us to investigate the role of non-Markovian noise played in the resource cost of PEC. Hakoshima *et al.*[51] demonstrated that non-Markovianity can reduce the PEC resource cost.

5. Mitigating measurement error

Since the rate of measurement errors is larger than that of two-qubit gate errors in some systems,[4,10] it is crucial to mitigate measurement errors. The widely discussed method to mitigate measurement errors is post-processing according to the results of quantum detector tomography.[52–54]

We consider a positive-operator valued measurement $\boldsymbol{M} = (M_1, M_2, \ldots, M_L)$. According to the Born's rule, the error-free probability of yielding M_i is $f_C^{M_i} = \mathrm{Tr}(M_i \rho^{\mathrm{ef}})$. The probabilities consist the vector $f_C^{\boldsymbol{M}} = (f_C^{M_1}, f_C^{M_2}, \ldots, f_C^{M_L})^{\mathrm{T}}$. We assume that the measurement error can be modeled as the matrix D which transforms the probabilities to $y_C^{\boldsymbol{M}}$, i.e.,

$$y_C^{\boldsymbol{M}} = D f_C^{\boldsymbol{M}}. \tag{34}$$

Elements of D are positive and satisfy $\sum_i D_{ij} = 1$, $\forall\ j$. D can be determined in experiments via quantum detector tomography. We consider that the error-free result f_C of our observable is a function of $f_C^{\boldsymbol{M}}$, i.e.,

$$f_C = G(f_C^{\boldsymbol{M}}), \tag{35}$$

where G denotes the function. For example, if the observable is the Pauli operator σ_z on the first qubit, the result is $f_C = f_C^{M_{|0\rangle}} - f_C^{M_{|1\rangle}}$, where $M_{|0\rangle}$ ($M_{|1\rangle}$) projects the first qubit onto $|0\rangle$ ($|1\rangle$). Combine Eqs. (34) and (35), the formula to mitigate measurement errors is

$$f_C' = G(D^{-1} y_C^{\boldsymbol{M}}). \tag{36}$$

In practice, due to the existence of coherent errors, $D^{-1} y_C^{\boldsymbol{M}}$ may yield un-physical results such as negative probabilities. In that case, we should replace D^{-1} in Eq. (36) with D' determined by

$$\arg\min_{D'} ||D' y_C^{\boldsymbol{M}} - D^{-1} y_C^{\boldsymbol{M}}|| \tag{37}$$

subjected to that $D' y_C^{\boldsymbol{M}}$ are normalized positive probabilities and $||\cdot||$ is the Euclidean norm.

Without multi-qubit correlations, D can be represented as a tensor product of single-qubit measurement errors, i.e.,

$$D = \bigotimes_{i=1}^{n} D^{(i)}, \tag{38}$$

where $D^{(i)}$ describes the measurement error on the i-th qubit. When there are correlated errors, we should employ efficient methods featuring these correlations.[55,56]

6. Virtual distillation

Recently, various virtual distillation (VD) methods were proposed in several works. We categorize these methods into two kinds: the multi-copy approach[22,23] and the state verification approach.[24,25]

Given an erroneous state ρ, the purified state $\rho^M/\mathrm{Tr}(\rho^M)$ has exponentially reduced errors. Suppose we have the following spectrum decomposition of the erroneous state:

$$\rho = P_0 |\psi_0\rangle\langle\psi_0| + \sum_{j=1}^{2^n-1} P_j |\psi_j\rangle\langle\psi_j|, \tag{39}$$

where $P_0, P_1, \ldots, P_{2^n-1}$ are in descending order and $P_0 - P_1 > 0$. VD can exponentially suppress error when $|\psi_0\rangle\langle\psi_0| = \rho^{\mathrm{ef}}$. This is because in the purified state

$$\frac{\rho^M}{\mathrm{Tr}(\rho^M)} = \frac{|\psi_0\rangle\langle\psi_0| + \sum_{j=1}^{2^n-1} (P_j/P_0)^M |\psi_j\rangle\langle\psi_j|}{1 + \sum_{j=1}^{2^n-1} (P_j/P_0)^M}, \tag{40}$$

the probability of $|\psi_0\rangle\langle\psi_0|$ exponentially approaches to 1 as M increases. Using the method proposed in the early works,[57,58] it requires $nM + \ln(M!)$ qubits to prepare this purified state. VD refers to those recently proposed methods to directly obtain the purified expectation

$$f'(C) = \frac{\mathrm{Tr}(O\rho_C^M)}{\mathrm{Tr}(\rho_C^M)} \tag{41}$$

without actually preparing the purified state, and in contrast, VD requires at most $nM + 1$ qubits.

VD estimates the numerator and denominator in Eq. (41) separately. The multi-copy approach utilizes

$$\mathrm{Tr}(O\rho_C^M) = \mathrm{Tr}\left(O S^{(M)} \rho_C^{\otimes M} \right), \tag{42}$$

where the observable O is on one of the copies and $S^{(M)}$ is the cyclic permutation operator, i.e., $S^{(M)}|\psi_1\rangle|\psi_2\rangle\cdots|\psi_M\rangle = |\psi_2\rangle|\psi_3\rangle\cdots|\psi_1\rangle$. We can verify Eq. (42) by explicitly writing both the left- and right-hand sides into summations of matrix entries, while here we provide an intuitive interpretation given by Koczor.[22] Suppose there are $M = 2$ copies of states prepared as $|\psi\rangle|\psi\rangle$. If any one of the two copies is flipped (due to an error) to an orthogonal state $|\tilde{\psi}\rangle$, measuring the operator $OS^{(2)}$ will yield $\langle\psi|\langle\tilde{\psi}|OS^{(2)}|\psi\rangle|\tilde{\psi}\rangle = \langle\psi|O|\tilde{\psi}\rangle\langle\tilde{\psi}|\psi\rangle = 0$, thus the contribution of the error vanishes in the purified expectation. To compute the right-hand side of Eq. (42), we can prepare M copies of the erroneous state and measure the expectation of $OS^{(M)}$. To measure the expectation, one approach is measurement by diagonalization,[23] i.e., we can append extra gates to the end of the original circuit to transform $OS^{(M)}$ into a sum of measurable observables. The transformation procedure is practical when the number of copies M is small, since $S^{(M)}$ can be represented as tensor products of qubit-wise permutation operators which lies in 2^M-dimensional Hilbert space. The other way to measure the expectation of $OS^{(M)}$ is by utilizing the Hadamard-test technique.[22,59] We here provide an example circuit[22] in Fig. 2

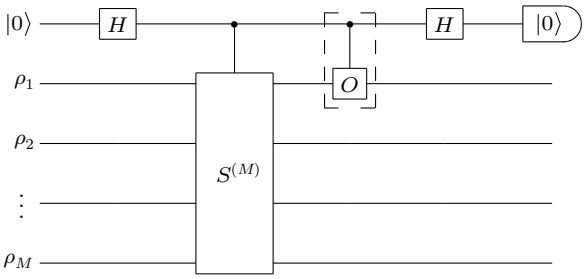


Fig. 2. An example circuit to implement VD via the Hadamard-test technique. The qubit on the top is the ancilla. M copies of quantum states $\rho_1, \rho_2, \ldots, \rho_M$ are prepared in parallel on the left. The controlled-$S^{(M)}$ gate can be decomposed into $M-1$ controlled-SWAP between states. The numerator in Eq. (41) is computed as $\mathrm{Tr}(O\rho_C^M) = 2P(0) - 1$, where $P(0)$ is the probability that the output state of the ancilla is $|0\rangle$. The denominator in Eq. (41) is computed as $\mathrm{Tr}(\rho_C^M) = 2P'(0) - 1$, where $P'(0)$ is the probability without the controlled-O gate in the dashed box.

The state verification approach is to check whether or not the dual form of the state-preparation circuit can evolve the state back to its initial. We call this procedure the dual test. In practical quantum computation, the expectation of an observable is measured by averaging over many circuit executions. State verification only keeps those executions that pass the dual test and discards the rest. The state-verified circuit usually has an intermediate measurement sandwiched between the state-preparation circuit and its dual circuit. The intermediate measurement can be executed in a non-destructive manner with the aid of one auxiliary qubit. We here provide an example circuit[25] in Fig. 3. The multi-copy approach and state-verification approach can be combined[60,61] to achieve qubit overhead lower than $nM + 1$.

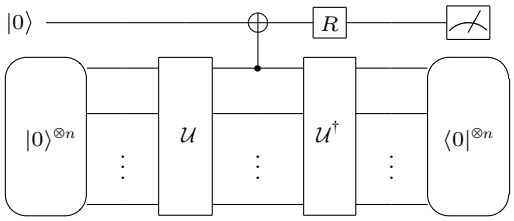


Fig. 3. An example of state verification circuit. The circuit is used to compute $f' = \mathrm{Tr}(\sigma_z\rho^2)/\mathrm{Tr}(\rho^2)$, where $\rho = \mathcal{U}(|0\rangle\langle 0|^{\otimes n})$ and σ_z is the Pauli-Z operator on the first qubit of ρ, i.e., the control qubit of the CNOT gate in this figure. The qubit on the top is the ancilla and the Pauli-Z (-X) operator on the ancilla is denoted by $\sigma_z^{(a)}$ ($\sigma_x^{(a)}$). R can be either I or H. When $R = I$, we measure $\langle\sigma_z^{(a)}\rangle$ of the ancilla. When $R = H$, we can effectively measure $\langle\sigma_x^{(a)}\rangle$ of the ancilla by measuring in the Pauli-Z basis. The distilled result is $f' = \langle\sigma_z^{(a)}\rangle_0/(1 + \langle\sigma_x^{(a)}\rangle_0)$, where $\langle O\rangle_0$ denotes the expectation of O conditioned on the output state being $|0\rangle^{\otimes n}$.

Here we briefly discuss the efficacy of VD. The variance is increased by VD due to the denominator $\mathrm{Tr}(\rho^M) < 1$. Additionally, extra circuit executions are required in practical implementations. Firstly, measurement by diagonalization may map $OS^{(M)}$ to the sum of several observable that can not be measured at the same time. We may have to estimate each observable individually and thus more circuit executions are needed. Secondly, state verification requires more circuit executions due to the discarding of failed trials in the dual test. The bias in VD comes from two aspects. One is due to the finite number of M, however this part of bias can be exponentially suppressed by increasing M. The other one is due to the so-called coherent mismatch[22,62] or the noise floor,[23] i.e., the difference between ρ^{ef} and $|\psi_0\rangle\langle\psi_0|$: $\Delta = 1 - \mathrm{Tr}(\rho^{\mathrm{ef}}|\psi_0\rangle\langle\psi_0|)$. Suppose that $\rho = (1-\epsilon)\rho^{\mathrm{ef}} + \epsilon\rho_{\mathrm{e}}$, the coherent mismatch vanishes only if $[\rho^{\mathrm{ef}}, \rho_{\mathrm{e}}] = 0$. Given in Refs. [23,62], for general states and observables, the bias caused by coherent mismatch is bounded by

$$|f_C - f_C'| \leq 2\sqrt{\Delta}||O||_\infty, \tag{43}$$

where $||O||_\infty$ is the largest absolute eigenvalue of O.

7. Symmetry verification and subspace expansion

Symmetry verification and subspace expansion are used to mitigate errors in the simulation of quantum systems. When the simulated system exhibits certain symmetries, we can mitigate errors by verifying these symmetries.[20,21] In quantum subspace expansion,[26,27] we can express the Hamiltonian in an effective subspace spanned by a set of noisy quantum states and obtain the error mitigated ground energy via generalized matrix diagonalization.

7.1. Symmetry verification

A symmetry is a unitary operator S that commutes with the system's Hamiltonian H, i.e.,

$$[H, S] = 0. \tag{44}$$

Suppose P_S is the operator that projects into the subspace of symmetry S, i.e., the subspace with the right eigenvalue of S. By measuring and post-selecting those states that pass the symmetry check of S,[20,21] we can project the erroneous state ρ onto

$$\rho' = \frac{P_S \rho P_S}{\mathrm{Tr}(P_S \rho P_S)}. \tag{45}$$

Since ρ' has excluded the states not lying in the subspace, it has lower error comparing with ρ. Accordingly, the error mitigation formula is

$$f_C' = \frac{\mathrm{Tr}(O P_S \rho P_S)}{\mathrm{Tr}(P_S \rho P_S)}, \tag{46}$$

i.e., $f_C' = y_{C_1}/y_{C_2}$ with $y_{C_1} = \mathrm{Tr}(O P_S \rho P_S)$ and $y_{C_2} = \mathrm{Tr}(P_S \rho P_S)$. In the case of $S^2 = \mathbb{I}$, the symmetry operator satisfies $S|\psi\rangle = s|\psi\rangle$, where $|\psi\rangle$ is an arbitrary state in the subspace and $s = \pm 1$. The corresponding projector is in the form of $P_S = (\mathbb{I} + sS)/2$. For example, the parity of the total electron number N_e is a symmetry if the system has a conserved number of electrons. The parity check operator P_{N_e} of the total electron number satisfies $P_{N_e}|\psi\rangle = s|\psi\rangle$ for all physical states of this system and $s = \pm 1$ depending on the parity of N_e. We can use the circuit in Fig. 4 to check the parity of the electron number. In the case that the measurement operator O commutes with the symmetry operator, we can directly measure O without the symmetry checking circuit and post-selecting those results that have the right symmetry.

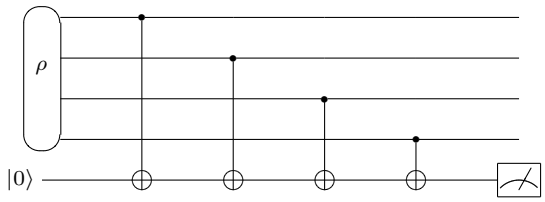


Fig. 4. A symmetry verification circuit for parity check. The qubit at the bottom is the ancilla. If spin-orbitals are mapped to qubits with $|1\rangle$ ($|0\rangle$) representing occupied (unoccupied) by an electron, we can check the parity of electron numbers by measuring the ancilla.

7.2. Quantum subspace expansion

We briefly introduce quantum subspace expansion with the example of an application in solving electronic structure problems.[26,27] Suppose that the quantum circuit C can prepare the ground state $|\psi_g^{\mathrm{ef}}\rangle$ on an ideal device, and on a noisy device the prepared state is $|\psi_g\rangle$. We choose a set of expansion operator $\{S_0 = \mathbb{I}, S_1, S_2, \ldots\}$ that excites $|\psi_g^{\mathrm{ef}}\rangle$ to states with higher energies. S_i can be chosen as $S_i = a_\alpha^\dagger a_\beta$ where $a_\alpha^\dagger, a_\beta$ are fermion creation and annihilation operators, which are encoded as Pauli operators on a quantum computer with methods such as Jordan–Wigner transformation. To implement the subspace expansion method, we first measure the quantities $A_{ij} = \langle\psi_g|S_i^\dagger H S_j|\psi_g\rangle$ and $B_{ij} = \langle\psi_g|S_i^\dagger S_j|\psi_g\rangle$ and carry out the generalized matrix diagonalization

$$A\Xi = B\Xi\Lambda, \tag{47}$$

where A and B are matrices with respective entries A_{ij} and B_{ij}, $\Lambda = \mathrm{diag}(\Lambda_1, \Lambda_2, \ldots, \Lambda_i, \ldots)$ is a diagonal matrix of eigenvalues Λ_i, and Ξ is the matrix of eigenvectors. The error mitigation formula for the ground state energy is

$$f_C' = \min\{\Lambda_i | A\Xi = B\Xi\Lambda\}. \tag{48}$$

If we choose the set of expansion operators as $\{\mathbb{I}, S\}$, quantum subspace expansion and symmetry verification are equivalent.[21] Let $A_{ij} = \mathrm{Tr}(\rho S_i^\dagger H S_j)$ and $B_{ij} = \mathrm{Tr}(\rho S_i^\dagger S_j)$. Using the expansion set $\{\mathbb{I}, S\}$, solving Eq. (47) is equivalent to diagonalizing

$$B^{-1}A = \frac{1}{1 - \mathrm{Tr}(S\rho)^2}\begin{pmatrix} \alpha & \beta \\ \beta & \alpha \end{pmatrix}, \tag{49}$$

where

$$\alpha = \mathrm{Tr}(H\rho) - \mathrm{Tr}(S\rho)\mathrm{Tr}(HS\rho), \tag{50}$$

$$\beta = \mathrm{Tr}(HS\rho) - \mathrm{Tr}(H\rho)\mathrm{Tr}(S\rho). \tag{51}$$

We note that B is invertible if $\mathrm{Tr}(S\rho) \neq 1$. The eigenvalues of Eq. (49) are

$$\Lambda_i = \frac{\mathrm{Tr}(H\rho) \pm \mathrm{Tr}(HS\rho)}{1 \pm \mathrm{Tr}(S\rho)}. \tag{52}$$

The error mitigated expectation of H according to the symmetry-verification formula in Eq. (46) is

$$\mathrm{Tr}(H\rho') = \frac{\mathrm{Tr}(H P_S \rho P_S)}{\mathrm{Tr}(P_S \rho P_S)}. \tag{53}$$

Substituting $P_s = (\mathbb{I} + sS)/2$ into Eq. (53) and considering $[H, P_S] = 0$, we have

$$\mathrm{Tr}(H\rho') = \frac{\mathrm{Tr}(H\rho) + s\mathrm{Tr}(HS\rho)}{1 + s\mathrm{Tr}(S\rho)}, \tag{54}$$

which admits the same form as that in Eq. (52).

7.3. Symmetry expansion

Symmetry expansion[63,64] can be regarded as the general form of symmetry verification and subspace expansion using symmetry operators for expansion. The complete symmetry group of an error-free quantum state is the stabilizer group $\mathbb{S}$. Define the projector

$$P_{\mathbb{S}'} = \frac{1}{|\mathbb{S}'|}\sum_{S \in \mathbb{S}'} S, \tag{55}$$

where $\mathbb{S}' \subseteq \mathbb{S}$ is a subset of the whole stabilizer group. If we choose $\mathbb{S}' = \{\mathbb{I}\}$, $P_{\mathbb{S}'}$ means no projection and error is not mitigated. If we choose $\mathbb{S}' = \mathbb{S}$, the noisy state is projected onto the intersections of all subspaces of symmetries which contain only the error-free state, and the error-free result is obtained. The practical choice of $\mathbb{S}'$ in the NISQ era, is to let $\mathbb{S}'$ be the intersection between the Pauli group and the stabilizer group, hence symmetry expansion is reduced to verify the Pauli operator symmetries.

8. Error mitigation using Clifford-circuit learning

The Clifford-circuit learning method was proposed by Strikis *et al.*[18] and Czarnik *et al.*[19] Generally speaking, given an error mitigation formula $f'_C = F(y_{C_1}, y_{C_2}, \ldots, \lambda_1, \lambda_2, \ldots)$, Clifford circuit learning obtains the parameters by minimizing the loss function

$$\langle (f'_C - f_C)^2 \rangle_{C\in\mathbb{C}}, \tag{56}$$

where $\langle\cdot\rangle_{C\in\mathbb{R}}$ denotes average over the set $\mathbb{R}$, and $\mathbb{C}$ are training circuits generated by randomly replacing all non-Clifford single-qubit gates in a circuit with Clifford gates. Here we have assumed that all multi-qubit gates are all Clifford. In Eq. (56), $f'_{C'}$ can be computed on the noisy quantum computer according to the formula F and $f_{C'}$ can be computed on a classical computer utilizing the efficient simulation of Clifford circuits.[65]

In principle, we can take any form for F. Strikis *et al.*[18] took the form of PEC in Eq. (19) and addressed two issues in PEC. The first issue is, PEC according to gate-wise decomposition, cannot completely eliminate (both spacial and temporal) correlated errors among gates. The second issue is that the decomposition we find using the method in Eq. (24) is optimal for every gate, however, may not be optimal for the whole circuit. In the learning-based protocol, we can directly optimize the quasi-probability decomposition *ab initio*, and we take into account the joint quasi-probability among gates rather than the products of gate-wise quasi-probabilities. Therefore the two issues are addressed. It is demonstrated that the learning-based PEC is advantageous in both reducing the bias and variance comparing with the gate-wise PEC when correlated errors are significant.

Alternatively, Czarnik *et al.*[19] used the error-mitigation formula with two factor a and b:

$$f'_C = a y_C + b. \tag{57}$$

The corresponding loss function is $\langle (a y_C + b - f_C)^2 \rangle_{C\in\mathbb{C}}$. When $b = 0$, Eq. (57) is also the formula of error mitigation with the phenomenological error model[66] and with the global depolarizing noise.[67–69]

8.1. The efficacy of Clifford learning

The efficacy of Clifford learning lies upon two results. This first result is that we can evaluate quantum-circuit error loss by sampling Clifford circuits[70] when the error model is independent of single-qubit gates. The underlying reason is that Clifford circuits formulate a unitary-2 design,[71] hence we have

$$\langle (f'_C - f_C)^2 \rangle_{C\in\mathbb{C}} = \langle (f'_C - f_C)^2 \rangle_{C\in\mathbb{U}}, \tag{58}$$

where $\mathbb{U}$ is the set of circuits when single-qubit gates are arbitrary unitary. When the error model depends on single-qubit gates, Eq. (58) does not hold in general. However we can still obtain a reliable estimation of the loss over unitary circuits using the method of hybrid sampling,[70] where a few gates in the circuit can be general unitaries while the other gates are all Clifford.

The second result is that, the bias can be complete eliminated in the ideal case such as when learning-based PEC finds the optimal quasi-probabilities.[18] Even for simple formulas such as that in Eq. (57), the bias is sufficiently reduced.[66,72] The results in Ref. [66] suggest that, if f'_C is optimized using Clifford circuits, it holds that

$$\frac{\sqrt{\langle (f'_C - f_C)^2 \rangle_{C\in\mathbb{U}}}}{\sqrt{\langle (y_C - f_C)^2 \rangle_{C\in\mathbb{U}}}} \propto \frac{1}{\sqrt{N}}, \tag{59}$$

where N is the number of gates.

Sampling Clifford circuits is efficient if we use the importance Clifford sampling technique.[66] Suppose that the observable is a Pauli operator, the error-free result of a Clifford circuit is either 0 or ± 1. These Clifford circuits whose error-free results are zero do not contribute to the loss in Eq. (56), because their erroneous results are also zero when subjected to Pauli errors. For example, ρ^{ef} is an error-free final state of a Clifford circuit with $\mathrm{Tr}(O\rho^{\mathrm{ef}}) = 0$. When subjected to a depolarizing channel, the erroneous final state is $\rho = (1-\alpha)\rho^{\mathrm{ef}} + \alpha\mathbb{I}$, where α is the depolarizing rate, whose expected output is also zero. Importance Clifford sampling efficiently produces Clifford circuits with ± 1 outcomes, thus the loss can be efficiently estimated.

9. Combined methods

Multiple error mitigation methods can be used in combination when quantum computing with one method individually can not achieve a satisfactory accuracy. We can in principle take the output of one error-mitigated method as the input of another. For any two error mitigation formulas F_1 and F_2, the concatenated formula is

$$f''_C = F_2(f'_{C_1}, f'_{C_2}, \ldots, f'_{C_i}, \ldots), \tag{60}$$

where f'_{C_i} is the error mitigated result according to the formula F_1. We can also concatenate multiple formulas in the same way. Next, we briefly introduce the combined methods proposed in literature.

Symmetry verification can be used in combination with error extrapolation and probabilistic error cancellation.[32] Symmetry verification has its limitations since those errors that commute with the symmetry operator can not be mitigated. Thus, it is natural to use other methods to mitigate the remaining errors. For example, we can replace $y_C(\epsilon_j)$ in Eq. (3) with the symmetry verified result $y^S_C(\epsilon_j)$ according to

the formula in Eq. (46) and the error mitigation formula becomes

$$f'_C = \sum_{j=0}^{m} \lambda_j y_C^S(\epsilon_j). \tag{61}$$

Probabilistic error cancellation and error extrapolation can also be used in combination.[32,73] We can use error extrapolation to determine the gate-wise decomposition in Eq. (20) for noise-agnostic gates, i.e., we can replace $\mathcal{U}_{i,\beta_i}$ with the error-boosted gate $\mathcal{U}_i(\epsilon^{\beta_i})$ and determine the quasi-probabilities λ_{i,β_i} using the error extrapolation method. On the other hand, if we have the tomographic knowledge on the noise operations, we can use PEC to effectively decrease the noise rate. We can use the results at several different decreased noise rates as the data points in error extrapolation. For example, in Eq. (3), ϵ_j can be considered as the j-th error rate reduced using PEC.

The framework of generalized subspace expansion[74] enables us to take the outputs of error extrapolation and virtual distillation as the inputs of subspace expansion. The general framework permits us to write the entries of A in Eq. (47) as $A_{ij} = \mathrm{Tr}(\rho^{(i)}\rho^{(j)}H)$. We can let $\rho^{(i)}$ be either the output of virtual distillation with i copies, or the output of i-th error-boosted quantum circuit.

All kinds of methods can be used in combination with Clifford-circuit learning.[66,75] For example, given a two-points error extrapolation formula

$$f'_C = (1-\lambda)y_C(\epsilon_1) + \lambda y_C(\epsilon_2), \tag{62}$$

we can use the Clifford-circuit learning method to determine the parameter λ. Another example is, we can always take the error mitigated result as the input of Clifford circuit learning. Supposing that f'_C is the error-mitigated result via PEC or virtual distillation, we can further improve the result by $f''_C = af'_C$, where a is determined using Clifford-circuit learning.

10. Summary

In this paper, we provide a thorough overview of the state-of-the-art QEM methods in a general form, where each method is expressed as an error mitigation formula. With the formula, we can concatenate any two or more QEM methods by replacing the input y to the high-level method with output f' of the low-level method. The formula can be further optimized with lower bias and variance using the Clifford-circuit learning method, where the formula is written as a parameterized function.

Note that the qubit cost and implementation complexity are not reflected in the error mitigation formula. The efficiency of some methods, for instances PEC and Clifford circuit sampling, is maximized under the assumption of sampling random circuits. Random circuit sampling is necessary in Pauli twirling to convert general errors into Pauli errors, which have favorable properties in QEM. In this regard, some QEM methods are easier to implement. For example, as error extrapolation requires no random circuit sampling, it can be implemented with only a few circuits.

The bias and the variance of the QEM estimator depend on the error mitigation formula and estimator of each y_C. In general, there is room to optimize y_C estimators to maximize the performance of QEM. For example, in PEC, circuit executions are allocated according to importance sampling. With the QEM formula, we can analyze the bias and the variance using the standard approach, e.g., propagation of uncertainty.

Error mitigation methods are important for quantum computing on NISQ devices. Compared with QEC, QEM can work in quantum circuits with high error rates. On the other hand, as the variance increases with the circuit size, more sampling and circuit executions are required to reduce the variance, thus QEM is not as scalable as QEC. However, as the QEM technique continuously evolves, we anticipate some benchmarking tools to be developed to reduce the circuit execution overhead. Meanwhile, QEM could be used in part of QEC to reduce the qubit cost when fault-tolerant technology becomes available.[76–79]

Acknowledgments

This work is supported by the National Natural Science Foundation of China (Grant Nos. 11875050 and 12088101) and NSAF (Grant No. U1930403).

References

[1] Nielsen M A and Chuang I L 2010 *Quantum Computation and Quantum Information* (Cambridge University Press) pp. 171–276
[2] Ladd T D, Jelezko F, Laflamme R, Nakamura Y, Monroe C and O'Brien J L 2010 *Nature* **464** 45
[3] Bao F, Deng H and Ding D, *et al.* 2021 arXiv:2111.13504 [quant-ph]
[4] Arute F, Arya K and Babbush R, *et al.* 2019 *Nature* **574** 505
[5] Kim Y, Wood C J, Yoder T J, Merkel S T, Gambetta J M, Temme K and Kandala A 2021 arXiv:2108.09197 [cond-mat, physics:quant-ph]
[6] Terhal B M 2015 *Rev. Mod. Phys.* **87** 307
[7] Babbush R, Gidney C, Berry D W, Wiebe N, McClean J, Paler A, Fowler A and Neven H 2018 *Phys. Rev. X* **8** 041015
[8] Preskill J 2018 *Quantum* **2** 79
[9] Shor P W 1996 *Proceedings of 37th conference on foundations of computer science*, October 14–16, 1996, Burlington, VT, USA, p. 56
[10] Wu Y, Bao W S and Cao S, *et al.* 2021 *Phys. Rev. Lett.* **127** 180501
[11] McArdle S, Endo S, Aspuru-Guzik A, Benjamin S C and Yuan X 2020 *Rev. Mod. Phys.* **92** 015003
[12] Bharti K, Cervera-Lierta A, Kyaw T H, Haug T, Alperin-Lea S, Anand A, Degroote M, Heimonen H, Kottmann J S, Menke T, Mok W K, Sim S, Kwek L C and Aspuru-Guzik A 2022 *Rev. Mod. Phys.* **94** 015004
[13] Li Y and Benjamin S C 2017 *Phys. Rev. X* **7** 021050
[14] Temme K, Bravyi S and Gambetta J M 2017 *Phys. Rev. Lett.* **119** 180509
[15] Endo S, Benjamin S C and Li Y 2018 *Phys. Rev. X* **8** 031027
[16] Magesan E, Gambetta J M and Emerson J 2011 *Phys. Rev. Lett.* **106** 180504
[17] Greenbaum D 2015 arXiv:1509.02921 [quant-ph]
[18] Strikis A, Qin D, Chen Y, Benjamin S C and Li Y 2021 *PRX Quantum* **2** 040330

[19] Czarnik P, Arrasmith A, Coles P J and Cincio L 2020 arXiv:2005.10189 [quant-ph]
[20] McArdle S, Yuan X and Benjamin S 2019 *Phys. Rev. Lett.* **122** 180501
[21] Bonet-Monroig X, Sagastizabal R, Singh M and O'Brien T E 2018 *Phys. Rev. A* **98** 062339
[22] Koczor B 2021 *Phys. Rev. X* **11** 031057
[23] Huggins W J, McArdle S, O'Brien T E, Lee J, Rubin N C, Boixo S, Whaley K B, Babbush R and McClean J R 2021 *Phys. Rev. X* **11** 041036
[24] O'Brien T E, Polla S, Rubin N C, Huggins W J, McArdle S, Boixo S, McClean J R and Babbush R 2021 *PRX Quantum* **2** 020317
[25] Huo M and Li Y 2022 *Phys. Rev. A* **105** 022427
[26] McClean J R, Schwartz M E, Carter J and de Jong W A 2017 *Phys. Rev. A* **95** 042308
[27] Colless J, Ramasesh V, Dahlen D, Blok M, Kimchi-Schwartz M, McClean J, Carter J, de Jong W and Siddiqi I 2018 *Phys. Rev. X* **8** 011021
[28] Rubin N C, Babbush R and McClean J 2018 *New J. Phys.* **20** 053020
[29] Peruzzo A, McClean J, Shadbolt P, Yung M H, Zhou X Q, Love P J, Aspuru-Guzik A and O'Brien J L 2014 *Nat. Commun.* **5** 4213
[30] Wang K, Chen Y A and Wang X 2021 arXiv:2111.00691 [math-ph, physics:quant-ph]
[31] Giurgica-Tiron T, Hindy Y, LaRose R, Mari A and Zeng W J 2020 *IEEE International Conference on Quantum Computing and Engineering (QCE)*, October 12–16, 2020, Denver, CO, USA, pp. 306–316
[32] Cai Z 2021 *npj Quantum Inf.* **7** 80
[33] Wang P, Luan C Y, Qiao M, Um M, Zhang J, Wang Y, Yuan X, Gu M, Zhang J and Kim K 2021 *Nat. Commun.* **12** 233
[34] Krebsbach M, Trauzettel B and Calzona A 2022 arXiv:2201.08080 [cond-mat, physics:quant-ph]
[35] Bishop C 2006 *Pattern Recognition and Machine Learning* (Springer) pp. 4–11
[36] Wallman J J and Emerson J 2016 *Phys. Rev. A* **94** 052325
[37] Dumitrescu E F, McCaskey A J, Hagen G, Jansen G R, Morris T D, Papenbrock T, Pooser R C, Dean D J and Lougovski P 2018 *Phys. Rev. Lett.* **120** 210501
[38] Klco N, Dumitrescu E F, McCaskey A J, Morris T D, Pooser R C, Sanz M, Solano E, Lougovski P and Savage M J 2018 *Phys. Rev. A* **98** 032331
[39] He A, Nachman B, de Jong W A and Bauer C W 2020 *Phys. Rev. A* **102** 012426
[40] Schultz K, LaRose R, Mari A, Quiroz G, Shammah N, Clader B D and Zeng W J 2022 arXiv:2201.11792 [quant-ph]
[41] Garmon J W O, Pooser R C and Dumitrescu E F 2020 *Phys. Rev. A* **101** 042308
[42] Kandala A, Temme K, Córcoles A D, Mezzacapo A, Chow J M and Gambetta J M 2019 *Nature* **567** 491
[43] McKay D C, Alexander T, Bello L, Biercuk M J, Bishop L, Chen J, Chow J M, Córcoles A D, Egger D, Filipp S, Gomez J, Hush M, Javadi-Abhari A, Moreda D, Nation P, Paulovicks B, Winston E, Wood C J, Wootton J and Gambetta J M 2018 arXiv:1809.03452 [quant-ph]
[44] https://quanlse.baidu.com
[45] Otten M and Gray S K 2019 *Phys. Rev. A* **99** 012338
[46] Otten M and Gray S K 2019 *npj Quantum Inf.* **5** 11
[47] Takagi R 2021 *Phys. Rev. Research* **3** 033178
[48] Piveteau C, Sutter D and Woerner S 2022 *npj Quantum Inf.* **8** 12
[49] Jiang J, Wang K and Wang X 2021 *Quantum* **5** 600
[50] Sun J, Yuan X, Tsunoda T, Vedral V, Benjamin S C and Endo S 2021 *Phys. Rev. Applied* **15** 034026
[51] Hakoshima H, Matsuzaki Y and Endo S 2021 *Phys. Rev. A* **103** 012611
[52] Chen Y, Farahzad M, Yoo S and Wei T C 2019 *Phys. Rev. A* **100** 052315
[53] Kwon H and Bae J 2021 *IEEE Transactions on Computers* **70** 1401
[54] Maciejewski F B, Zimboás Z and Oszmaniec M 2020 *Quantum* **4** 257
[55] Geller M R and Sun M 2021 *Quantum Sci. Technol.* **6** 025009
[56] Bravyi S, Sheldon S, Kandala A, Mckay D C and Gambetta J M 2021 *Phys. Rev. A* **103** 042605
[57] Berthiaume A, Deutsch D and Jozsa R 1994 *Proceedings Workshop on Physics and Computation*. PhysComp'94, November 17–20, 1994, Dallas, TX, USA, pp. 60–62
[58] Barenco A, Berthiaume A, Deutsch D, Ekert A, Jozsa R and Macchiavello C 1997 *SIAM J. Comput.* **26** 1541
[59] Cotler J, Choi S, Lukin A, Gharibyan H, Grover T, Tai M E, Rispoli M, Schittko R, Preiss P M, Kaufman A M, Greiner M, Pichler H and Hayden P 2019 *Phys. Rev. X* **9** 031013
[60] Cai Z 2021 arXiv:2107.07279 [quant-ph]
[61] Czarnik P, Arrasmith A, Cincio L and Coles P J 2021 arXiv:2102.06056 [quant-ph]
[62] Koczor B 2021 *New J. Phys.* **23** 123047
[63] McClean J R, Jiang Z, Rubin N C, Babbush R and Neven H 2020 *Nat. Commun.* **11** 636
[64] Cai Z 2021 *Quantum* **5** 548
[65] Aaronson S and Gottesman D 2004 *Phys. Rev. A* **70** 052328
[66] Qin D, Chen Y and Li Y 2021 arXiv:2112.06255 [quant-ph]
[67] Vovrosh J, Khosla K E, Greenaway S, Self C, Kim M S and Knolle J 2021 *Phys. Rev. E* **104** 035309
[68] Urbanek M, Nachman B, Pascuzzi V R, He A, Bauer C W and de Jong W A 2021 *Phys. Rev. Lett.* **127** 270502
[69] Wang X, Feng X, Funcke L, Hartung T, Jansen K, Kühn S, Polykratis G and Stornati P 2021 arXiv:2111.15522 [hep-lat, physics:quant-ph]
[70] Wang Z, Chen Y, Song Z, Qin D, Li H, Guo Q, Wang H, Song C and Li Y 2021 *Phys. Rev. Lett.* **126** 080501
[71] Dankert C, Cleve R, Emerson J and Livine E 2009 *Phys. Rev. A* **80** 012304
[72] Dalzell A M, Hunter-Jones N and Brandão F G S L 2021 arXiv:2111.14907 [quant-ph]
[73] Mari A, Shammah N and Zeng W J 2021 *Phys. Rev. A* **104** 052607
[74] Yoshioka N, Hakoshima H, Matsuzaki Y, Tokunaga Y, Suzuki Y and Endo S 2022 *Phys. Rev. Lett.* **129** 020502
[75] Bultrini D, Gordon M H, Czarnik P, Arrasmith A, Coles P J and Cincio L 2021 arXiv:2107.13470 [quant-ph]
[76] Xiong Y, Chandra D, Ng S X and Hanzo L 2020 *IEEE Access* **8** 228967
[77] Piveteau C, Sutter D, Bravyi S, Gambetta J M and Temme K 2021 *Phys. Rev. Lett.* **127** 200505
[78] Lostaglio M and Ciani A 2021 *Phys. Rev. Lett.* **127** 200506
[79] Suzuki Y, Endo S, Fujii K and Tokunaga Y 2022 *PRX Quantum* **3** 010345